Structure vs métrique dans les graphes

David Coudert, Guillaume Ducoffe, Nicolas Nisse

Structure vs métrique dans les graphes †

David Coudert[1,2] and Guillaume Ducoffe[2,1] and Nicolas Nisse[1,2]

[1] *Inria, France*
[2] *Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France*

L'émergence de réseaux de très grande taille oblige à repenser de nombreux problèmes sur les graphes : en apparence simples, mais pour lesquels les algorithmes de résolution connus ne passent plus à l'échelle. Une approche possible est de mieux comprendre les propriétés de ces réseaux complexes, et d'en déduire de nouvelles méthodes plus efficaces. C'est dans ce but que nous démontrons des relations générales entre les propriétés *structurelles* des graphes et leurs propriétés *métriques*. Nos relations se déduisent de nouvelles bornes serrées sur le diamètre des *séparateurs minimaux* dans un graphe.

Plus précisément, nous prouvons que dans tout graphe G le diamètre d'un séparateur minimal S dans G est au plus $\lfloor \ell(G)/2 \rfloor \cdot (|S|-1)$, avec $\ell(G)$ la plus grande taille d'un cycle isométrique dans G. Nos preuves reposent sur des propriétés de connexité dans les *puissances* d'un graphe. Une conséquence de nos résultats est que pour tout graphe G, sa longueur arborescente (*treelength*) est au plus $\lfloor \ell(G)/2 \rfloor$ fois sa largeur arborescente (*treewidth*). En complément de cette relation, nous bornons la largeur arborescente par une fonction de la longueur arborescente et du *genre* du graphe. Cette borne se généralise à la famille des graphes qui excluent un *apex-graph H* comme mineur. Par conséquent, nous obtenons un algorithme très simple qui, étant donné un graphe excluant un apex-graph fixé comme mineur, calcule sa largeur arborescente en temps $O(n^2)$ et avec facteur d'approximation $O(\ell(G))$.

Keywords: Graphe; Largeur arborescente (Treewidth); Longueur arborescente (Treelength); Hyperbolicité; Genre

1 Introduction

La croissance de la taille des réseaux (par exemple Twitter ou Facebook) rend désormais indispensable la conception d'algorithmes rapides pour obtenir des solutions exactes. Des questions *a priori* simples comme calculer le diamètre peuvent devenir difficiles et requièrent de nouvelles techniques [CGH+13]. Ainsi la tendance actuelle est de mieux comprendre les propriétés de ces réseaux complexes pour les utiliser à des fins algorithmiques.

Il est bien connu que de nombreux problèmes peuvent être résolus efficacement dans des graphes dont la structure est proche de celle des arbres. Plus précisément, de nombreux problèmes NP-complets en général peuvent être résolus en temps polynomial dans la classe des graphes de *largeur arborescente* (treewidth) bornée [Cou90]. Malheureusement, deux inconvénients majeurs font que cet outil théorique très puissant ne peut être utilisé en pratique. D'une part, ces algorithmes prennent en entrée une décomposition arborescente des graphes qui doit donc être calculée. Or, calculer la treewidth est NP-difficile [ACP87] et les algorithmes d'approximation existants reposent sur une complexe programmation semi-définie [FHL08]. D'autre part, de nombreux réseaux réels (comme celui des systèmes autonomes (AS) de l'Internet) ont une grande largeur arborescente, i.e., leur structure est très éloignée de celle d'un arbre [dMSV11].

Ce second fait a motivé une autre approche, qui consiste à étudier les propriétés métriques des réseaux plutôt que leurs propriétés structurelles. En effet, des expériences suggèrent que les réseaux complexes, par exemple des réseaux biologiques ou sociaux, sont métriquement proches des arbres [AAD14, ADM14]. Un paramètre qui mesure la proximité de la métrique d'un graphe (sa distribution des distances) à celle d'un arbre est son *hyperbolicité* au sens de Gromov [Gro87]. Ce paramètre a été très étudié pour ses

† Ce travail est soutenu par l'ANR, projets Stint ANR-13-BS02-0007 et "Investissements pour l'Avenir" ANR-11-LABX-0031-01, l'équipe partenaire INRIA AlDyNet et le projet ECOS-Sud Chile. Les preuves omises sont accessibles dans [CDN14].

avantages algorithmiques [KL06], principalement pour des problèmes de routage [BPK10]. L'hyperbolicité d'un graphe est liée à une autre mesure des décompositions arborescentes, à savoir, la *longueur arborescente* (treelength) [DG07]. La longueur arborescente n'a pas, semble-t-il les mêmes avantages algorithmiques que la largeur arborescente, cependant elle a des intérêts non négligeables. D'une part, bien que son calcul soit NP-complet [Lok10], il existe des algorithmes très efficaces (basés sur des BFS) qui, étant donné un graphe $G = (V,E)$, calculent une décomposition arborescente de G de longueur au plus trois fois l'optimal en temps $O(|V| \cdot |E|)$ [DG07]. D'autre part, la longueur arborescente des réseaux complexes est faible, comme par exemple celle du graphe des AS [dMSV11].

Notre but est donc de tirer parti à la fois des avantages de la longueur arborescente (calculable efficacement, faible dans les grands réseaux réels) et de ceux de la largeur arborescente (existence d'algorithmes efficaces dans les graphes de largeur bornée). Pour cela nous établissons de nouveaux liens entre ces paramètres, ce qui nous permet notamment d'obtenir un algorithme efficace pour le calcul approché de la largeur arborescente dans une large classe de graphes. Nos résultats sont en fait plus généraux puisqu'ils considèrent les séparateurs minimaux des graphes.

Définitions. Un graphe connexe $G = (V,E)$ est dit *triangulé* si tous ses cycles induits (sans corde) sont de longueur trois. Une *triangulation* de G est un graphe triangulé obtenu de G en ajoutant des arêtes. Une *décomposition arborescente* de G est équivalente à une triangulation T de G. La largeur d'une décomposition est la taille d'une clique maximum. Sa longueur est la distance maximum dans G entre deux sommets adjacents dans H. La *largeur arborescente* de G, notée $tw(G)$ (treewidth), est la largeur minimum d'une décomposition de G. Sa *longueur arborescente*, notée $tl(G)$ (treelength), est la longueur minimum d'une décomposition de G. Finalement, un sous-graphe H de G est *isométrique* si la distance entre toute paire de sommets de H est la même dans H que dans G. Notons que $\ell(G)$, la longueur d'un plus long cycle isométrique de G, peut être calculée en temps polynomial [Lok09].

Pour finir, rappelons qu'un *séparateur minimal* $S \subseteq V$ de G est un ensemble de sommets tel qu'il existe deux composantes connexes A et B de $G \setminus S$ dans lesquelles chaque sommet de S a au moins un voisin. Le diamètre de S est la distance maximum dans G entre deux sommets de S.

Premières remarques. Pour tout graphe G, $tl(G)$ est une borne supérieure de son hyperbolicité et est majorée par son diamètre [DG07]. De plus, $tl(G)$ est au plus la moitié de la longueur d'un plus grand cycle induit de G [DG07]. Cependant, longueur et largeur arborescentes sont incomparables en général, ce qui écarte la possibilité d'une relation simple entre les deux invariants. Par exemple, pour le cycle C_n à n sommets, $tl(C_n) = \lceil \frac{n}{3} \rceil$ et $tw(C_n) = 3$, alors que pour la clique K_n à n sommets, $tw(K_n) = n-1$ et $tl(K_n) = 1$. L'exemple de C_n laisse penser que la présence de long cycles isométriques est un facteur pour qu'un graphe ait une grande longueur arborescente. En particulier, la longueur arborescente d'un graphe G est au moins $\lceil \ell(G)/3 \rceil$ [DG07]. Cependant, cela n'est pas suffisant puisque les grilles ne contiennent pas de cycles isométriques de taille 5 ou plus, mais ont une longueur arborescente arbitrairement grande. A l'inverse de la clique, pour tout graphe planaire G, $tw(G) \leq 12 \cdot tl(G)$ [DG09]. Il est naturel de se demander si l'écart entre largeur et longueur arborescentes est borné par une fonction du *genre* du graphe.

Contributions. Nous introduisons une nouvelle méthode, très générique, pour borner le diamètre d'un séparateur minimal dans un graphe G en fonction de la taille du séparateur et de propriétés globales de G. En particulier, nous démontrons que pour tout séparateur minimal S de G, le diamètre de S est au plus $\lfloor \ell(G)/2 \rfloor \cdot (|S|-1)$. Nous déduisons de notre méthode que, si G n'est pas un arbre, alors on a $tl(G) \leq \lfloor \ell(G)/2 \rfloor \cdot (tw(G)-1)$[‡].

Par ailleurs, nous prouvons que $tw(G) = O\left(g\sqrt{g}\right) \cdot tl(G)$ pour tout graphe G de genre g, ce que nous généralisons à tous les graphes excluant un *apex-graph* H fixé comme mineur[§].

En combinant ces résultats, nous obtenons que *tout* algorithme pour calculer la longueur arborescente

[‡] Récemment, et indépendamment de ce travail, Diestel et Müller ont prouvé que $tl(G) \leq \ell(G) \cdot (tw(G)-1)$ [DM14]. Notre borne est plus fine que la leur, et notre résultat plus général puisqu'il s'applique à tous les séparateurs minimaux du graphe.

[§] Un *apex-graph* est un graphe obtenu d'un graphe planaire en lui ajoutant un sommet dont le voisinage n'est pas contraint.

peut être transformé en un algorithme approché pour la largeur arborescente. Plus précisément, pour tout graphe G à n sommets et de genre g, les algorithmes de [DG07] calculent une $O\left(\ell(G)\cdot g\sqrt{g}\right)$-approximation de la largeur arborescente en temps quadratique $O(n^2)$.

2 Diamètre des séparateurs minimaux

Dans cette section, nous montrons que dans tout graphe G, le diamètre de tout séparateur minimal S est borné par le produit de sa taille et de $\lfloor \ell(G)/2 \rfloor$. Intuitivement, si deux sommets x et y de S sont dans la même composante connexe du sous-graphe induit par S, leur distance est au plus $|S|-1$. Sinon, nous prouvons qu'il existe un chemin de x à y dans G passant consécutivement par des composantes connexes $C_1,\cdots,C_k$ de S et tel que : $x\in C_1$, $y\in C_k$ et, pour $1\le i<k$, il existe un cycle isométrique de G qui traverse C_i et C_{i+1} ; deux composantes C_i,C_{i+1} consécutives sont donc à distance au plus $\lfloor \ell(G)/2 \rfloor$.

Nous utilisons la notion de *base des cycles* d'un graphe G, qui est un ensemble de cycles qui génère (par différence symmétrique des arêtes) tous les sous-graphes Eulériens de G. Il est connu qu'il existe toujours une base composée de cycles isométriques. En d'autres termes, tout cycle d'un graphe peut être obtenu comme la différence symmétrique d'un ensemble de cycles isométriques.

Théorème 1 *Pour tout graphe G, tout séparateur minimal S de G a diamètre au plus* $\lfloor \ell(G)/2 \rfloor\cdot(|S|-1)$

Idée de preuve. Nous prouvons en fait un résultat plus fort : pour tout graphe G qui possède une base des cycles dont tous les cycles sont de longueur au plus $l\ge 3$, tout séparateur minimal de G induit un sous-graphe connexe de $G^{\lfloor l/2 \rfloor}$, avec $G^{\lfloor l/2 \rfloor}$ le graphe obtenu de G en rendant adjacents toute paire de sommets qui sont à distance au plus $\lfloor l/2 \rfloor\ge 1$ dans G. Soit $\mathcal{G}_l$ l'ensemble des graphes G qui admettent une telle base des cycles.

Étant donnés $G\in\mathcal{G}_l$ et S un séparateur minimal de G, nous prouvons d'abord qu'il existe deux composantes de S qui sont traversées par un cycle de la base des cycles de G, i.e., il existe deux composantes de S qui sont à distance au plus $\lfloor l/2 \rfloor$ dans G (Propriété $(*)$). La propriété se démontre en considérant deux composantes connexes A et B de S. Puisque S est minimal, il existe un cycle C qui traverse A et B. Par définition, il existe un ensemble de cycles de taille au plus l qui génère C. Nous prouvons que parmi ces cycles, il en existe un qui traverse deux composantes connexes de S (pas nécessairement A et B).

Ensuite, nous transformons G en un graphe H, que nous obtenons en ajoutant des arêtes entre tous les sommets de S à distance au plus $\lfloor l/2 \rfloor$, puis en contractant en un sommet chaque composante de S dans H. Nous prouvons que $H\in\mathcal{G}_l$, et enfin que l'ensemble S' de sommets de H qui résulte de la contraction des composantes de S est aussi un séparateur minimal de H. Par l'absurde, si S' était composé de plusieurs composantes connexes, alors par construction elles devraient être deux-à-deux à distance au moins $\lfloor l/2 \rfloor+1$ dans G (puisque des arêtes sont ajoutées entre les composantes à distance au plus $\lfloor l/2 \rfloor$). Mais cela contredit la Propriété $(*)$. $\square$

L'exemple du cycle montre que cette borne est optimale en général. Nous améliorons cette borne dans certaines classes de graphes particulières [CDN14].

3 Relations entre longueur et largeur arborescentes

Dans cette section, nous prouvons que largeur et longueur arborescentes sont équivalentes dans la classe des graphes de genre et de plus long cycle isométrique bornés. Cela nous permet d'obtenir un nouvel algorithme d'approximation pratique pour le calcul de la largeur arborescente.

En utilisant les résultats de [BT01], une application directe du Théorème 1 permet de prouver que $tl(G)=O(tw(G))$ dans la classe des graphes de cycles isométriques bornés.

Théorème 2 *Pour tout graphe G qui n'est pas un arbre,* $tl(G)\le\lfloor \ell(G)/2 \rfloor\cdot(tw(G)-1)$.

Nous prouvons ensuite que la relation inverse, i.e., $tw(G)=O(tl(G))$, existe dans les graphes G de genre borné. Pour cela, nous utilisons des outils puissants de la théorie de la bi-dimensionalité [DHT06].

Intuitivement, pour tout graphe G de genre borné g, soit G a une petite largeur arborescente et la relation est vraie, soit G contient comme contraction un graphe H qui est obtenu d'une triangulation planaire d'une grille de côté $O(tw(G)/g)$ en ajoutant $O(g)$ arêtes. Comme la longueur arborescente est close par contraction, $tl(G) \geq tl(H)$. Nous prouvons le résultat en montrant que $tl(H) = \Omega\left(\frac{tw(H)}{g\sqrt{g}}\right)$.

Ce résultat peut être généralisé aux graphes excluant un *apex-graph* comme mineur [CDN14].

Théorème 3 *Soit H un apex-graph fixé. Pour tout graphe G excluant H comme mineur, il existe $c_H \in \mathbb{N}$ tel que $tw(G) \leq c_H \cdot tl(G)$. En particulier pour tout graphe G de genre $g > 0$, $tw(G) \leq 96 \cdot tl(G) \cdot g\sqrt{g}$.*

4 Conclusion

Nous déduisons des théorèmes 2 et 3 que pour tout graphe à n sommets et de genre g, l'algorithme de [DG07] calcule en temps $O(g \cdot n^2)$ un entier t^* tel que $tw(G)/(96 \cdot g\sqrt{g}) \leq t^* \leq 3\lfloor \ell(G)/2 \rfloor \cdot tw(G)$, ce qui fournit un nouvel algorithme d'approximation efficace de la largeur arborescente. Pour les graphes dont une borne supérieure sur la largeur arborescente est connue, l'algorithme fournit une borne inférieure sur le genre. À court terme, nous voudrions l'appliquer dans de grand réseaux réels pour en estimer leur largeur arborescente. Un inconvénient de l'algorithme est qu'il construit une décomposition de longueur t^* mais de largeur arbitraire. Notons qu'il existe des graphes G de n sommets pour lesquels toute décomposition arborescente de largeur $\alpha \cdot tw(G)$ et de longueur $\alpha \cdot tl(G)$ requiert $\alpha = \Omega(n^{1/5})$ [DG07]. Une étape importante sera de rendre notre approche constructive. Peut-on convertir cette décomposition, de longueur t^*, en une décomposition de largeur $O(\ell(G) \cdot g\sqrt{g} \cdot t^*)$ (quite à augmenter sa longueur) ?

Références

[AAD14] M. Abu-Ata and F. F. Dragan. Metric tree-like structures in real-life networks : an empirical study. Technical Report arXiv :1402.3364, ArXiv, 2014.

[ACP87] S. Arnborg, D. G. Corneil, and A. Proskurowski. Complexity of finding embeddings in a k-tree. *SIAM J. Algebraic Discrete Methods*, 8(2) :277–284, April 1987.

[ADM14] R. Albert, B. DasGupta, and N. Mobasheri. Topological implications of negative curvature for biological and social networks. *Physical Review E*, 89(3) :032811, 2014.

[BPK10] M. Boguna, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature Communications*, 1(62) :1–18, October 2010.

[BT01] V. Bouchitté and I. Todinca. Treewidth and minimum fill-in : Grouping the minimal separators. *SIAM J. Comput.*, 31(1) :212–232, 2001.

[CDN14] D. Coudert, G. Ducoffe, and N. Nisse. Diameter of minimal separators in graphs. Technical Report RR-8639, Inria Sophia Antipolis ; I3S, 2014. https ://hal.archives-ouvertes.fr/hal-01088423.

[CGH$^+$13] P. Crescenzi, R. Grossi, M. Habib, L. Lanzi, and A. Marino. On computing the diameter of real-world undirected graphs. *Theoretical Computer Science*, 514 :84–95, 2013.

[Cou90] B. Courcelle. The monadic second-order logic of graphs. i. recognizable sets of finite graphs. *Information and Computation*, 85(1) :12 – 75, 1990.

[DG07] Y. Dourisboure and C. Gavoille. Tree-decompositions with bags of small diameter. *Discrete Mathematics*, 307(16) :2008–2029, 2007.

[DG09] Y. Dieng and C. Gavoille. On the tree-width of planar graphs. *Electronic Notes in Discrete Maths*, 34 :593–596, 2009.

[DHT06] E. D. Demaine, M. Hajiaghayi, and D. M. Thilikos. The bidimensional theory of bounded-genus graphs. *SIAM J. Discrete Math.*, 20(2) :357–371, 2006.

[DM14] R. Diestel and M. Müller. Connected tree-width, nov 2014. http ://arxiv.org/abs/1211.7353.

[dMSV11] F. de Montgolfier, M. Soto, and L. Viennot. Treewidth and hyperbolicity of the internet. In *10th IEEE International Symposium on Network Computing and Applications (NCA)*, pages 25–32, Boston, 2011. IEEE.

[FHL08] U. Feige, M. Hajiaghayi, and J. R. Lee. Improved approximation algorithms for minimum weight vertex separators. *SIAM J. Comput.*, 38(2) :629–657, 2008.

[Gro87] M. Gromov. Hyperbolic groups. *Essays in Group Theory*, 8 :75–263, 1987.

[KL06] R. Krauthgamer and J.R. Lee. Algorithms on negatively curved spaces. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 119–132. IEEE, 2006.

[Lok09] D. Lokshtanov. Finding the longest isometric cycle in a graph. *Discrete Applied Maths*, 157(12) :2670–2674, 2009.

[Lok10] D. Lokshtanov. On the complexity of computing treelength. *Discrete Applied Mathematics*, 158(7) :820–827, 2010.